

# 統合検索のための 共通メタデータと歴博データベース のデータ項目のマッピング

Mapping of REKIHAKU Database's Fields to Common Metadata  
for Integrated Retrieval

安達文夫

ADACHI Fumio

はじめに

①統合検索システムの概要と対象データベース

②データ項目のマッピング

③マッピングの評価と考察

むすび

## 【論文要旨】

人間文化研究機構では、機構を構成する各機関が公開する100を超える多種多様なデータベースをまとめて検索できる統合検索システムを構築し、2008年4月よりサービスを開始した。このシステムは、個々のデータベースの所在や操作方法を意識することなく検索できることに特徴がある。多くのデータベースと同様に、検索語とのマッチング、検索結果の一覧表示と詳細表示を基本機能としているが、個別に作成されたデータ項目がまちまちなデータベースをまとめて検索する上で、共通メタデータが主要な役割を果たしている。この共通メタデータに、個々のデータベースのデータ項目をいかにマッピングするかが重要な課題であった。統合検索の対象とした国立歴史民俗博物館の43のデータベースをマッピングする上での考え方を記した。

見やすい詳細表示とするには、共通メタデータの特定の要素に、データ項目が集中しないようマッピングする必要がある。この方針に沿って、共通メタデータとして採用したDC (Dublin Coreメタデータ) に、実物の資料を記述したデータベースを適用することは、要素の意味の拡張が必要となって、適合性が低い。このことを、博物館資料の情報流通向けに提案した博物館コアメタデータの要素とDCの要素の対応付けを通して明確にした。一方、博物館コアメタデータは、資料に関するデータベースだけでなく、遺跡や種々の文献から研究の関心を抽出した事項、ならびに研究文献に関するデータベースにおいても、そのマッピングに親和性がある。これは、記述の要素として役割や出来事の内容を問わない形で人物、時間、空間を取り出すとともに、共通性が高く記述対象を特徴づける名称、種別、主題を要素としていることが大きい。より汎用的なメタデータを求めるなら、残りの要素の抽象化を図るとともに、共通化の可能性を探ることが重要となる。

【キーワード】 横断検索, 資源共有, ダブリンコアメタデータ, 博物館資料, 人文科学情報

---

## はじめに

人間文化研究機構では、研究資源共有化事業として、機構を構成する各機関が公開する総計で100を超えるデータベースをまとめて検索できる統合検索システムを構築し [1]、第Ⅰ期のシステムとして、2008年4月よりサービスを開始した。その後、機能拡充を行い、2012年より第Ⅱ期のシステムによる検索サービスを提供している。国立歴史民俗博物館（以下、歴博と記す）としては、2015年4月時点で43件のデータベースを、同システムより公開している。

統合検索システムは、個々のデータベースの所在や操作方法を意識することなく検索できることに大きな特徴がある。また、時間や場所の情報を基に検索したり、検索結果を地図上や年表上に表示する機能を有している。このシステムを実現する上での主要な課題の一つは、個別に作成されたデータ項目がまちまちなデータベースをまとめて検索するために設ける共通のメタデータこれを共通メタデータと呼ぶを如何に構成するかということにあった。個々のデータベースの共通メタデータに即したメタデータを、個別のレコード毎に作成するのは膨大な手数がかかり実際的ではない。そこで、データベースのデータ項目とメタデータの要素との対応付けのルールを定め、これに基づいて機械的にマッピングを行う。このデータベース毎のマッピングのルールを、どのように定めるかが重要な課題となった。

この統合検索システムを構築する以前より、人文科学に関するデータベースを横断的に検索する方法の検討とそのプロトタイプシステムの構築が進められてきた [2]。ここでは、汎用的なメタデータとして有望であった Dublin Core Metadata (DC メタデータ、あるいは DC と記す) [3] が共通メタデータとして採用された。歴博から筆者らもその構築に加わり、歴博の館蔵資料 DB と館蔵中世古文書 DB を具体的な対象として、これらのデータ項目の DC へのマッピングについて検討を行った [4]。その結果、一応のマッピングはできるものの、DC が元来ネットワーク上の資源の記述を目的としていたことに起因して、資料の持つ物理的な属性に関する対応付けの適合性が低いことが明らかとなった。

統合検索システムにおいても、DC を共通メタデータとすることで開発を進めた。そこで、歴博のデータベースを統合検索の対象とすることを念頭に、館蔵資料に関するデータベースの対象を広げるとともに、研究文献データベースや研究成果データベースをも対象として、DC のドキュメント [3, 5, 6] に基づいて、データ項目を対応付けるための DC の要素の意味の合理的な拡張について検討を進めた [7-9]。そして、博物館資料の情報の流通に適する博物館コアメタデータの提案を行い [10]、統合検索システムの機能拡充の際に、共通メタデータの役割の見直しを行って、一部のデータベースに対して、このメタデータの適用を図った。

DC メタデータは、統合検索の第Ⅰ期のシステムにおいて、共通メタデータとして中心的な役割を果たし、第Ⅱ期のシステムにおいても同等の役割を有している。このため、歴博のデータベースのデータ項目を DC にマッピングする際の考え方を記録として残すことに意味がある。そして、このマッピングの問題点を明らかにしておくことは、今後より汎用的な共通メタデータを検討する上で重要と考えられる。

そこで、本論では、①で共通メタデータに関する事柄を中心に、統合検索システムの概要を記し、②において、データベースの種類ごとに、データ項目のDCへのマッピングとその適合性を記す。所蔵資料に関するデータベースについては、博物館コアメタデータの要素を導出したモデルを示し、この要素とDCとの対応付けを論ずる。③において、実際のDCへのマッピングの評価を示すとともに、汎用的な共通メタデータへの考察を加える。

## ① 統合検索システムの概要と対象データベース

### 1.1 機能の概要

統合検索システムの検索とその結果の表示に関する基本的なフローを図1に示す。多くのデータベースが普通に有する検索語入力と検索結果の一覧表示ならびに詳細表示を基本とする。詳細表示は統合検索の検索結果の統一的なビューを与えるが、元のデータベースの詳細のすべてではない。そこで、元のデータベースの検索結果の詳細画面を表示する機能を設けている。また、複数のデータベースをまとめて検索することで、ヒット件数がかんりの数になることから、一覧表示するデータベースを選択できるヒット件数表示の画面を設けている。

検索語入力に関して、第Ⅰ期のシステムでは、検索の分解能が異なる全項目検索、人・物・時・所検索、名称・作者・主題検索、詳細検索およびDC検索の5種類の手段を提供した。利用時の分かりづらさがあったことから、第Ⅱ期のシステムでは、全項目検索を簡易検索と呼ぶこととし、人・物・時・所検索と名称・作者・主題検索を合せたものを詳細検索として、2つの手段に整理して提供している。

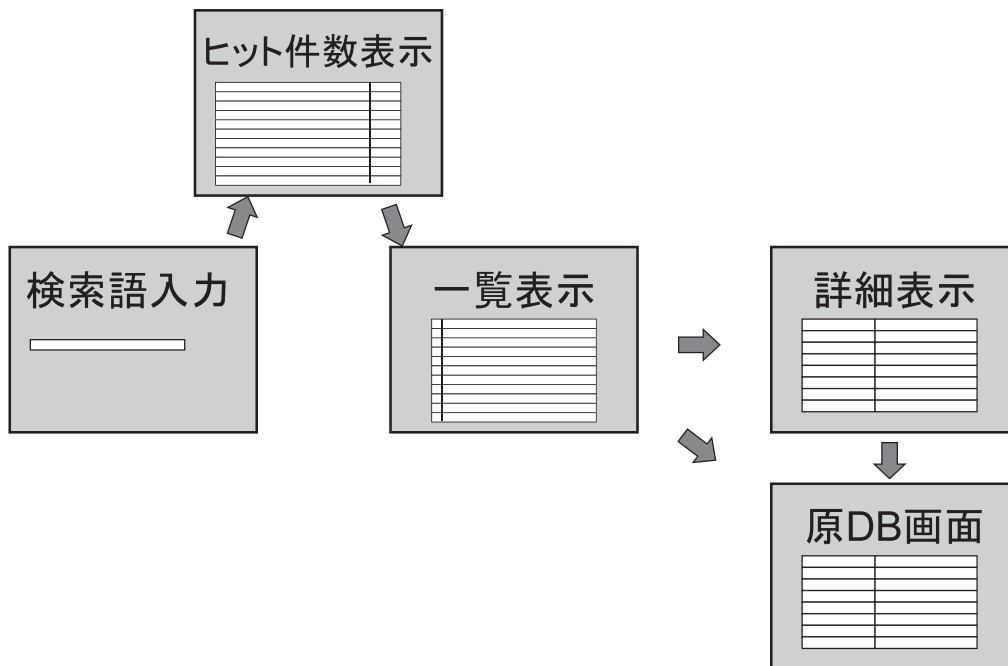


図1 統合検索の基本フロー

## 1.2 共通メタデータ

統合検索において共通メタデータは、検索の対象となるとともに、一覧表示および詳細表示の際の表示用文字データとしての役割を果たす。このとき、法量のような数字で記述された項目の検索は意味がないことから、検索の対象外となるよう、共通メタデータを検索用と表示用に分けて設けている。また、第Ⅰ期と第Ⅱ期のシステムで共通メタデータの一部を変えている。これらのメタデータの関係を図2に示す。

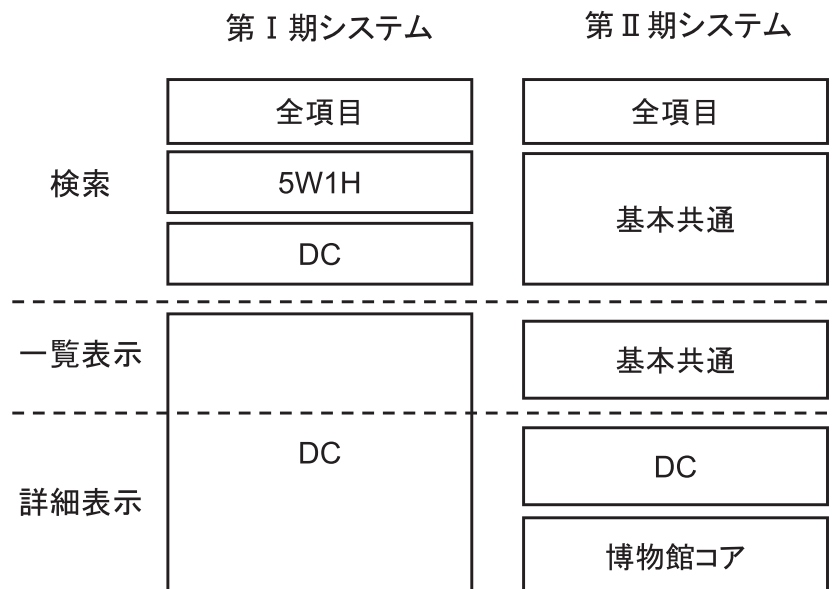


図2 統合検索における共通メタデータ統合検索における共通メタデータ

第Ⅰ期のシステムでは、共通メタデータとして検索用、表示用とも、比較的広く使用されているDCを基本とした。DCの基本要素は15の項目からなる。この中のCoverageは、対象とする時間的・空間的範囲を記すもので、時間と空間の記述の混在が避けられない。これらを分けた検索および表示が必要であることから、Coverageのサブ要素であるTemporalとSpatialを使用することとした。この一部修正したDCを表1に示す。

このDCに加えて、全項目検索用に全項目メタデータ、および人・物・時・所検索用に5W1Hメタデータと呼ぶ、Who, What, When, Where, Othersを要素とする共通メタデータを用意した。5W1Hメタデータを設けたのは、例えば人物データベースの人に関する項目が、DCの人に関する要素であるCreator, Contributor, Publisherにマッピングされるとは限らず、DCから5W1Hメタデータを生成できないことによる。全項目メタデータは、DCあるいは5W1Hメタデータの全ての要素を検索対象とすればあらためて設ける必要はない。これを設けたのは実装上の事情による。

検索結果の表示には、検索語入力の手段によらずDCメタデータを適用した。

第Ⅱ期のシステムでは、検索対象として全ての要素を選択できるDC検索を廃止し、検索手段として簡易検索と詳細検索の2つに絞るとともに、検索結果の表示において、一覧表示の項目はデータベースに依らず共通とするが、詳細表示は幾つかの形態があってよいとした。これに伴い、共通

表1 統合検索システムに適用した DC の構成

要素	意味
Identifier	識別番号
Title	名称
Type	資源の性質や種類
Subject	主題
Creator	作成に主たる責任を持った個人・団体
Contributor	作成に寄与した個人・団体
Publisher	資源を利用可能にしたことに主たる責任を持つ個人・団体
Date	資源の事象に関する時間的情報
Temporal	時間的範囲
Spatial	空間的範囲
Format	物理形式またはデジタル化形式
Description	説明や記述
Relation	関連する資源
Source	資源が作り出される元となった資源
Rights	権利に関する情報
Language	記述している言語

メタデータとして検索用と一覧表示用に基本共通メタデータ [11] と称するものを適用し、詳細表示用には個々のデータベースの性質に合わせて上記の DC か随時定義するメタデータを適用できるようにした。

基本共通メタデータは、識別番号、名称、人物、時間、空間、キーワードの6個の要素からなる。5W1H メタデータと比べて、What に対応する要素がないのは、これに対するマッピングが分かりづらい点を改善したことによる。また、基本共通メタデータは others の要素を持たないため、第Ⅱ期のシステムでは、全項目メタデータが必須となる。

詳細表示用の随時定義するメタデータとして、博物館コアメタデータを用意した。このメタデータを構成する際の考え方について2.2節で触れる。

以上のように、第Ⅱ期のシステムから、DC は統合検索システムの中の全てもデータベースの基本的な共通メタデータではなくなった。しかしながら、人間文化研究機構の統合検索システムから見て外部の機関と連携した検索を行うには、DC を介することが現実的である。このため、全てのデータベースに対して、DC にマッピングしたメタデータを設けることを必須としている。

### 1.3 統合検索の対象データベース

歴博がネットワーク上で公開しているデータベースの中で、作成者の了解が得られたものを、統合検索の対象とした。このとき、歴博のデータベースと統合検索のデータベースを、当然ながら1対1に対応させることを基本とした。しかしながら、歴博のデータベースとして見掛けは一つであ

るが、研究の関心となる資料や遺跡とそれに関係する文献のように二つ以上の研究資源を同時に扱うものがある。一方、統合検索では、検索および検索結果の表示を研究資源毎に行う必要がある。そこで、統合検索では以下のように複数のデータベースとして扱った。

複数の研究資源を同時に扱う歴博のデータベースには、研究資源毎に検索語入力画面を設け、その検索結果表示画面からリンクにより関連する他の研究資源を表示する形式のデータベースと、複数の研究資源を一つの検索語入力画面で扱い、検索結果も一つの画面上で表示する形式のデータベースがある。

前者の形式の東国板碑 DB や中世地方都市 DB などでは、各々の研究資源を主体と捉えてマッピングするとともに、統合検索上は個別のデータベースとした。弥生石器遺跡 DB もこの形式に属し、遺跡、図面、文献の研究資源を対象としている。しかし、文献に関する検索語入力画面を持たないため、遺跡と図面の二つの DB を統合検索の対象とした。

後者の形式のデータベースとして、陶磁器出土遺跡 DB と城館城下発掘 DB の二つがある。いずれも遺跡と文献を扱い、どちらの研究資源も、検索対象となる項目を持つ。統合検索の共通メタデータとした DC では、二つの資源の一方を主体とし、他方をその関連とみなすことになるため、例えば遺跡を主体としたデータベースだけを構成した場合、文献に関する検索が不自然なものとなる。そこで、遺跡と文献のそれぞれを主体とした2種類のマッピングを行い、統合検索上個別のデータベースとした。土偶 DB もこの形式に当たる。しかし、一つの土偶に対して複数の文献が対応しているため、文献を主体とするマッピングの作業が煩雑になることから、土偶を主体としてマッピングする一つのデータベースとした。

以上により、統合検索システムの対象としたデータベースを表2に示す。ここでは、統合検索システムの種類別のデータベースグループに対応させて示している。事項・ファクトデータベースについては、後述の検討の関係から、さらに資料、遺跡、事項に分けて示している。

表2 歴博の統合検索の対象データベース

区分	データベース名	
研究文献データベース	荘園関係文献目録, 自由民権運動研究文献目録, 中世制札(文献), 中世地方都市(文献), 陶磁器出土遺跡(文献), 近世窯業関係主要文献目録, 東国板碑(文献), 日本民俗学文献目録, 民俗誌, 宮座研究論文	
所蔵資料データベース	館蔵資料, 館蔵紀州徳川家伝来楽器, 館蔵武器器具 実物資料, 館蔵縄文時代遺物, 館蔵中世古文書, 館蔵近世・近代古文書, 館蔵高松宮家伝来禁裏本, 館蔵『懐溜諸屑』, 館蔵武器器具 文献史料, 館蔵錦絵, 館蔵野村正治郎衣装, 館蔵染色用型紙, 館蔵装身具	
事項・ファクトデータベース	資料	棟札, 中世制札(制札), 土偶, 弥生石器(図面), 東国板碑(板碑)
	遺跡	陶磁器出土遺跡, 近世窯業遺跡, 城館城下発掘, 弥生石器遺跡, 東国板碑(遺跡等), 縄文・弥生集落遺跡
	事項	日本荘園, 古代・中世都市生活史(物価), 中世地方都市(都市), 江戸商人・職人, 民俗語彙, 俗信, 文化財材料知識(色材)知識



## ② データ項目のマッピング

本章では、表2に示したデータベースの種別毎に、そのデータ項目のDCメタデータへのマッピングの考え方について記す。このとき、データ項目との対応を直接示すのではなく、データベース群のデータ項目について共通する性質を記述要素として取り出し、データ項目と記述要素との対応を示した上で、記述要素とDCの要素との対応関係を見てゆく。所蔵資料データベースでは、博物館コアメタデータが記述要素にあたる。

### 2.1 研究文献データベース

表2の研究文献データベースのデータ項目より、共通する性質を記述要素として抽出したものを、各データ項目と対応させて、表3に示す。この記述要素とDCの要素との対応について確認する。

文献番号は、個々のデータベース毎に各レコードに付けられた番号である。DCのIdentifierは、その資源を指し示すユニークな番号を意味することから、厳密には、機関名+データベース名+文献番号がIdentifierとなる。しかし、簡単のため、文献番号を単独でIdentifierと考える。

表3の文献名称、著者、発行機関、発行年が、それぞれDCのTitle, Creator, Publisher, Dateに対応することに議論の余地はない。

表3で文献種別としている自由民権運動研究文献目録DBの形態分類キーワードには、資料集、目録、書評等の文献の区分が入力されている。これはgenreにあたるから、文献種別はDCのTypeに該当する。

同表で内容分類としている自由民権運動関係文献目録DBの内容キーワードには、結社、政党、差別・人権などが、民俗誌DBの分類名には人生儀礼、信仰、口承文芸などが、入力されている。これらはDCのSubjectに当たると見てよい。近世窯業関係主要文献目録の当該窯も同文献のSubjectと見なせる。

文献に記載された事項を特定の観点より抽出された項目であって、次の対象地域と対象年代以外のものを記載事項としている。これらはDescriptionとするのが適切であろう。但し、荘園名や遺跡名のように、文献が対象とする研究資源そのものはSubjectとすることも考えられる。

文献の記載対象の成立時期などの対象年代をTemporal、文献が対象とする対象地域をSpatialとすることに議論の余地はない。

文献のその他の記載事項やデータベース作成者のコメントが記入されている備考をDescriptionとすることも問題なかろう。

文献が掲載されている雑誌や報告書の名称である掲載誌は、DCに直接当てはまる要素はない。Relationに対応させることは不自然ではないと考えられる。

関連情報としている中世地方都市(文献)DBの初出、初出年月もRelationとするのが適切であろう。

以上をまとめて、記述要素とDCの要素との対応関係を表4に示す。このように、研究文献DBでは、ほとんど問題なくDCへのマッピングが定まる。

表3 研究文献データベースのデータ項目と記述要素との対応 (1/2)

要素	荘園関係 文献目録	自由民権運動研究 文献目録	中世制札 (文献)	中世地方都市 (文献)	陶磁器出土遺跡 (文献)	近世窯業関係 主要文献目録
文献番号	論文コード		文献ID	コード	文献ID	文献番号
文献名称	文献名	タイトル	文献著者	タイトル	書名	著書・論文
文献種別		形態分類キーワード				
内容キーワード		内容キーワード				当該窯
著者	文献著者	著者・作者名	文献名	著者	陶磁器部著者	編著者
発行機関		出版・発行所		出版社	発行者	発行所
発行年	出典年月	発行年月日	刊行年	年月	発行年	発行年
対象年代					時代	
対象地域	関連国名	地域別キーワード	関連国名	都道府県, 国名, 都市名	所在地, 市町村コード, 北緯, 東経	都道府県
記載事項	関連荘園名				遺跡名, 遺跡の種類, 中国・朝鮮・日本陶磁	
備考	文献備考		文献備考		特記事項	
掲載誌	出典文献	掲載誌・紙	文献出典	所収誌	シリーズ名	掲載誌
関連情報				初出, 初出年月		

表3 研究文献データベースのデータ項目と記述要素との対応 (2/2)

要素	城館城下発掘 (文献)	東国板碑 (文献)	日本民俗学 文献目録	民俗誌	宮座研究論文
文献番号	文献ID	文献番号	ID	文献番号	
文献名称	書名	論文・題名, 副題名	論文名	書名	論文名
文献種別					
内容キーワード				分類名	
著者	執筆者	編・著者	著者名	著者	著者
発行機関	発行機関	発行機関		出版者	
発行年	発行年	発行年	発行年月日	発行年	西暦年
対象年代	時代				
対象地域	所在地, 緯度・経度, 市コード		調査地, 市町村コード	対象地名, 市町村コード	ブロック, 府県
記載事項	遺跡名, 文献記述, 添載情報, 城館情報, 現状, 城館発掘, 城下発掘, 支援情報, 遺構種類, 調査面積, 記入者				
備考	特記事項				その他
掲載誌	シリーズ名	シリーズ名	収録雑誌, 巻号, 頁数	版・頁	雑誌名, 号数
関連情報					



表4 各種データベースの記述要素と DC の要素の対応

DC	研究文献 DB	資料 DB (博物館コアメタデータ)	遺跡 DB	事項 DB
Identifier	文献番号	識別番号	番号	識別番号
Title	文献名称	名称	遺跡名称	名称
Type	文献種別	種別	遺跡種別	種別
Subject	内容キーワード	主題		主題
Creator	著者	人物		人物
Contributor				
Publisher	発行機関			
Date	発行年	時間	時期	時間
Temporal	対象年代			
Spatial	対象地域	空間	所在地	空間
Format		数量 大きさ・質量 形状 材質	数 遺跡構造	
Description	記載事項 備考	記述	遺構情報 立地 遺物情報 調査情報 備考	記述
Relation	掲載誌 関連情報	関連	関連文献・資料	関連

## 2.2 所蔵資料データベース

前節で記した研究文献と異なり、博物館資料の情報を DC にマッピングするには旨く適合しない面がある。実際、歴博の所蔵資料に関するデータベースを対象として、そのデータ項目を DC の定義に即してマッピングすることを検討したが、その意味をかなり拡張する必要があった [8, 9]。このことから、DC に代わる博物館資料向けの統合検索用メタデータとして、博物館コアメタデータを提案した [10]。その要素数は DC の基本要素の数と同程度とすることを方針とし、歴博と国立民族学博物館の所蔵資料に関するデータベースのデータ項目全体を見て一般化することで要素を定めているが、その根拠についてはあまり触れていない。

本節では、資料情報がどのように付与されているかの一種のモデルを提唱し、これを基に統合検索向けのメタデータを導出した多少の根拠を示す。その上で、博物館コアメタデータと DC との対応付けについて述べる。これを通して、所蔵資料データベースの項目を DC にどのようにマッピングしたかを示す。

### 2.2.1 博物館資料情報付与モデル

ここでは、所蔵資料データベースのデータ項目がどのように定められているかを見る。言い換えると、資料目録の情報がどのように付されているかを整理する。そして、これを資料情報の付与の

---

流れとしてモデル化する。

その際に、資料が内容を持つものと持たないものがあることに留意する。すなわち、文献資料は(多くの場合)紙でできた媒体の上に文字で記された内容が載っていると見ることができる。資料は媒体と内容を合せたものと捉える。この見方は、DCの以前の版にあった resource と content の見方に類似している。屏風や絵巻のような絵画資料も、描かれた図像を内容として有していると見ることができる。一方、石器や土器は内容を持たない資料と見てよい。ここでは、内容を持たない資料をもの資料とよび、内容を有する文献資料と絵画資料を合せて文献絵画資料と呼ぶことにする。衣裳のように通常はもの資料とされる資料であっても、それが持つ文様について項目を立てて記述している場合は、本論では文献絵画資料に区分する。

以下、資料に関する情報がどのように付与されるかを記す。

資料の情報として、まず資料名称が与えられる。もの資料の場合は資料の名称が、文献資料では内容に基づいた題名が付与される。絵画資料の場合、錦絵では描かれた内容に応じた題名が与えられ、屏風などでは内容に基づく題名と資料の名称がセットになって付与されることが多い。

次に、資料を管理するための番号が与えられる。

さらに、資料の物としての属性である数量、大きさ・質量、形状、材質が付与される。このうち数量は管理の面から重要であり、大きさもどこを計るかは資料により異なるが早い段階で与えられる。

資料の整理・調査により明らかになってくる制作、使用、出土、伝来など、資料に生ずる出来事に関する何らかの事項が記される。

資料の種別を与えようとすることが多い。資料を一見して分類できる場合もあるだろうし、十分な調査によって判別される場合もある。資料の内容に対して種別を与えることもある。

資料中の文字による記載の一部あるいは全てを目録に転記することがある。もの資料であっても資料に記された印章・銘記などを項目を設けて記述することがある。一方、資料あるいは内容について、その調査・研究により得られる所見や解釈を記すことが多い。描かれた図像を読み取っての解釈もその一つである。ここでは、資料中の記載の転記を内部記述、解釈による外から与えた記述を外部記述と区分しておく。多くの場合、備考が設けられる。ここには、様々な事項が記され一概には言えないが、外部記述であることが多い。

外部記述の一種ではあるが、文献絵画資料では、その内容を特徴的あるいは端的に表す語を与えようとしている。

この他、他の資源等との関わりを表す関係が付されるものもある。

以上、博物館資料への情報の付与について順に記したが、当然ながらこの順に進むわけではない。調査・研究の結果、後になって付与されるものも多い。

### 2.2.2 メタデータ要素の抽出

博物館資料に付与される情報の全体を見ると、資料が元々有している情報は、ものとしての特性である数量、大きさ・質量、形状、材質と内部記述だけである。これと管理上付与される番号を除けば、資料に付与される情報は、資料名称を含めて、資料を説明するための外部記述の一つと見る

ことができる。博物館資料に適した統合検索用のメタデータを構成するには、この広義の外部記述から如何に要素として抽出するかが課題となる。

このとき、先に記したように、要素の総数は DC と同程度とした上で、意味が明確で検索および検索結果の表示に有用であり、データベース間での共通性が高いものを選んでゆく。なお、DC との相互交換性を意識し、要素名は DC と共通化あるいは抽象化を図る。以下、要素の抽出について記す。

資料名称は、資料を特徴付けることから必須である。考古資料では、使用名称が同一のものが多く、識別のため資料番号も必須となる。要素の名称としては、それぞれ名称と識別番号とする。

資料あるいは内容の種別、および内容を特徴的に表す語は、検索の手がかりとして有用である。これらを種別および主題とする。

資料が有するものとしての4つの属性をそのまま要素とする。

資料に生ずる出来事とその事象毎に記述することは、共通性が低下し要素数が多くなるため適さない。博物館資料情報構造化モデルでは、出来事を記述するが、その行為者、時期、場所を要素として表している [12]。ここでは出来事の事象は度外視し、3つの要素で表す点を参考にして、人物、時間、空間を要素として与え、共通化を図る。人物には機関も含むものとする。内部記述あるいは外部記述の中にも人物、時間、空間に関わる記述がある。このときは、これらの要素に対応させる。内部記述と狭義の外部記述を検索する上で区別することに意味はないことから、この二つをまとめて記述とする。

他の資料等との関係を一つの要素とする。

以上より、識別番号、名称、種別、主題、人物、時間、空間、数量、大きさ・質量、形状、材質、記述、関連が博物館コアメタデータの要素となる。ただし、文献 [10] では、ここでの記述を記述とその他に分け、DC の Description の定義に準じ、説明的なものだけを記述とし、これ以外をその他としている。

### 2.2.3 抽出要素の評価

表2に示した所蔵資料データベースのデータ項目を博物館コアメタデータの要素に対応させて表5に示す。但し、前述の議論を明らかにするため、種別を資料種別と内容種別に、記述を内部記述と外部記述に分けて示している。後述する DC との対応の検討のため、人物を Creator, Contributor, Publisher および Description に、時間を Date と Temporal に分けて記載している。また、様々な資料を対象としている館蔵資料 DB を除き、もの資料と文献絵画資料に大別して示している。なお、データ項目の記載に当たっては、資料名称（漢字）と資料名称（カナ）のように、内容が同一のものは、資料名称とまとめて示している。

表5の中で、データ項目の名称だけでは意味が不明なものについて注記する。館蔵『懷溜諸屑』の分類 a は寄席、商業、災害など資料に記された内容の大分類であり、分類 b はその中の落語、薬、地震など内容の小分類を表す。これに対し、分類 c は、半札、引札、商標など資料の種別を表し、形態も原本か刷物かを示し、資料の種別に当たる。館蔵錦絵の主題分類には風景画、美人画などの内容に関する種別が記され、内容分類には江戸名所、近江八景などの主題が入力されている。形態

表5 所蔵資料データベースのデータ項目と博物館コアメタデータの要素との対応

メタデータ要素		もの資料データベース			文献絵画資料データベース		
		館蔵資料	館蔵武器器具 実物資料	館蔵紀州徳川家 伝来楽器	館蔵縄文時代 遺物	館蔵中世古文書	館蔵近世・ 近代古文書
識別番号		資料番号	資料番号	資料番号	資料番号	資料番号, 卷子・冊子番号, 文書番号	資料番号
名称		資料名称	資料名称	資料名	資料名称	文書名	資料名称
種別	資料種別	実物・模造	種別	楽器種	種類		
	内容種別						
主題							
人物	Creator			作者			
	Contributor			下絵等著者, 銘書著者			
	Publisher						
	Description					差出, 充所	差出, 宛名
時間	Date	製作年代	時代(和暦), 時代(西暦)	時代 和暦, 時代 西暦	時期, 型式	時代	
	Temporal					日付(和暦, 西暦)	年月日, 西暦
空間		使用地	制作地		遺跡名, 所在地, 旧地名		
数量		数量		数量		紙数	数
大きさ・質量		法量, 尺度	法量	法量(縦, 横, 高さ, 径)	法量, 重量	法量(縦, 横)	縦・横
形状						形式	形態
材質		材質	材質		材質		
記述	内部記述		銘文	銘文		本文書出, 本文書止, 端裏書	
	外部記述	区分 備考	流派, 備考	品質・形状および 加飾, 修復歴, 伝来, 付属品, 文書数, 文書, 備考	出土状況 備考	欠簡・断簡 備考	備考
関連		コレクション名	コレクション名, 関連文献史料ID		コレクション名, 図録写真, 図録図版	区分, 所収文書群	コレクション名

文献絵画資料データベース							メタデータ要素 の出現割合
館蔵高松宮家 伝来禁裏本	館蔵 『懐溜諸屑』	館蔵武器武具 文献史料	館蔵錦絵	館蔵野村 正治郎衣装	館蔵染色用 型紙	館蔵装身具	
資料番号	資料番号	資料番号	資料番号	資料番号	資料番号	資料番号	1.00
資料名称, 外題, 内題, その他の題, 別書名・合綴書名	資料名称	資料名称	資料名称	資料名称	資料名称	資料名	1.00
	形態, 分類 c			種類		種別	0.62
			主題分類				
	分類 a, 分類 b	流派	内容分類, 人名, その他固有 件名	モチーフ	モチーフ	意匠	0.67
著編者	作成	著者	画工名			作者	0.62
		訳者	彫師, 摺師				
刊写年代・刊写者		版元	版元名				
	宛名	宛名					
	年代, 西暦年月日	時代(和暦), 時代(西暦)	和暦, 西暦	時代(和暦), 時代(西暦)	時代	時代, 和暦, 西暦年, 世紀	1.00
時期							
			画工住所, 版元住所, 国名, 地名				0.31
員数, 紙数	点数	数量	員数			数量	0.69
法量	法量(縦, 横)	法量	法量	法量(縦, 横)	彫刻面法量 (送巾, 彫巾), 地紙面法量 (縦, 横)	法量(縦, 横, 高さ)	1.00
装訂		形態	判型, 形態		種類		0.46
料紙				材質		材質	0.46
表紙等, 奥書等, 本文, 界線・匡郭, 付訓, 印記		内容詳細	画工署名 改印		捺印, 彫刻, 墨書	銘文	1.00
巻数, 注記, 備考	位置, 備考 1, 備考 2	備考	彩色 備考	技法, 糸込み 地色, 伝来 備考	伝来, 技法, 地紙の特徴, 備考	技法 備考	
	コレクション名	原著, コレクション名	コレクション名			コレクション名	0.69

には縦絵、横絵が入力されており、大判、間判、細判が入力されている判型とともに、形状に当たる。

博物館コアメタデータの要素が、表5に示したデータベースに出現する割合を、同表の右欄に示す。ここで、主題に関する母数は、文献絵画資料データベースの総数としている。要素の出現割合より、次の特徴が見られる。識別番号と名称は全てのデータベースに付されている。種別、主題は6割以上のデータベースに付されている。種別では内容に対して付与されているものがあることが見られる。人物、時間、空間について見ると、空間が3割程度とやや低いが、この3要素で記述することは妥当と考えられる。特に時間は歴史資料を対象としていることから、全てのデータベースに付与されている点特徴的である。ものとしての特性を表す数量、大きさ・質量、形状、材質もそれなりの割合で与えられている。特に大きさ・質量は必ず付与されている。記述は全てに見られる。これは、記述以外の要素に当てはまらない項目が、外部記述となることが要因となっている。内部記述だけを見ても、6割以上のデータベースに付与されている。関連には、歴博のコレクションを基とする資料管理に基づいて付与されるコレクション名が該当することから、出現割合が高くなっているが、これ以外にも他の資料との関連を示す項目が見られる。

以上のように、これらの要素は、出現割合が高く、意味も明確であることから、博物館資料情報の統合検索用のメタデータとして妥当であると考えられる。

## 2.2.4 DC との対応関係

所蔵資料に関するデータベースのデータ項目のDCへのマッピングは、第I期のシステムの構築の際に行い、第II期のシステムにおいては、外部システムとの連携の際のメタデータとして使用している。このマッピングは個々のデータ項目をDCの要素に直接対応付けることで行ったが、ここでは、博物館コアメタデータとDCの要素間の対応付けを見ることにより、全体を通しての課題として整理する。以下において、DCMI (Dublin Core Metadata Initiative) によるDCの要素の定義 [3] を引用する。この定義は変更が重ねられているが、第I期システムを構築するさいに参考とした2006年8月28日版のものを示す。

### (1) 対応性のある要素

まず、対応関係に議論の余地があまりないものについて確認を行う。

識別番号は、文献データベースの文献番号と同様に、それだけでは一意とは言えないが、Identifierに対応付けるのが妥当である。

名称を Title とすることに関し、多くの場合題名 (title) が与えられている文献絵画資料では問題がない。一方、もの資料では名前 (name) が付与されていることから多少の違和感が残る。しかし、DCMI の Title の定義は "A name given to the resource." であり、これに対応させることが適切である。

表5で主題としたデータ項目は資料の内容の topic を示すものと言え、DCMI で "The topic of the content of the resource." と定義される Subject に対応する。

記述は Description に対応する。厳密には、資料中の記載を転記した内部記述はDCMIの定義である "An account of the content of the resource." から外れるとも考えられるが、広い意味での



account と見て、Description とすることが適切と言える。

表 5 で関連としたデータ項目は、関連する他の資源を示し、Relation に対応させることに問題はない。

## (2) 1 対 n 関係の要素

次に、博物館コアメタデータから DC への対応関係が 1 対 n となるものについて見てみる。表 5 に示したように、人物の DC との対応は、資源の作成に関係する Creator, Contributor, Publisher および内部記述に該当する Description に分かれる。時間も、資源の制作等に関する Date とコンテンツの時間的な範囲を示す Temporal に分かれる。これらについて、データ項目と DC の要素との対応に問題はない。但し、博物館コアメタデータから DC へ直接変換できず、元のデータ項目から直接 DC にマッピングするか、博物館コアメタデータに表 5 と同様なサブ要素を設けるなどの手段を取る必要がある。

## (3) 検討を要する要素

上記以外は、DC の定義に無条件に合致する要素はない。このとき、これらを Description とすることが一つの考え方である。しかし、この場合、Description に様々な項目が混在し、詳細表示での見通しが悪くなる。これを避けるため、DC の要素を分散させることを基本とする対応付けを検討する。

空間に関する表 5 のデータ項目の中で、館蔵錦絵 DB の国名、地名は錦絵に描かれた地域を表し、空間的範囲を意味する Spatial に対応する。これ以外は、資源の制作などに関わる項目で、時間での Date に相当する Place のような要素に該当するものである。DC ではこのような要素はなく、Spatial に対応付けるのが妥当となる。

種別に当たる表 5 のデータ項目の中で、内容種別とした館蔵錦絵 DB の主題分類は、Type の DCMI の定義である "The nature or genre of the content of the resource." に一致する。一方、資料種別には例えば、紀州徳川家伝来楽器 DB の楽器種では、龍笛、明笛、和琴、鞆鼓などが、館蔵縄文遺物 DB の種類では石器、土器などが入力されている。これらを genre と見るのは微妙であるが、Type に対応させるのが適切であろう。

ものとしての特性を表す数量、大きさ・質量、形状、材質、に関連する DC の要素として Format がある。この DCMI による定義は "The physical or digital manifestation of the resource." とあり、その注記として "Typically, Format may include the media-type or dimensions of the resource. <中略> Examples of dimensions include size and duration. <後略>" とあることから、大きさはこれに対応するとしてよい。質量はこの定義に当てはまらないが、大きさ・質量は Format に対応すると見る。

数量も dimension の意味が拡張されたものと捉えれば Format に対応する。

形状は表 5 の中では文献絵画資料に見られる。先の記した館蔵錦絵の他には、冊子、卷子、折本、袋綴、堅紙などの語が入力されている。これらを physical manifestation と捉え、形状を Format に対応させる。ただし、最新の定義では、"The file format, physical medium, or dimensions of

---

the resource."とあり、形状はFormatに該当しなくなる。

材質に関し、Formatのサブ要素であるMediumの定義が"The material or physical carrier of the resource."とある。これを拠り所に、材質をFormatに対応させる。

このように、DCへの対応付けは、ものとしての特性を表す物理的な属性に対して適合性が低く、資料種別に対しても適切には合致しないことが確認される。

以上の博物館コアメタデータとDCの要素間の対応関係をまとめたものを、表4に示している。これと表5とを組み合わせることにより、それぞれのデータベースのデータ項目を、どのようにDCにマッピングしたかを見ることができる。

## 2.3 事項・ファクトデータベース

本節では、表2の事項・ファクトデータベースのデータ項目のDCメタデータへの対応付けについて、データベースの記述対象の区分である資料、遺跡、事項毎に、それぞれ記す。

### 2.3.1 資料

表2の資料に区分したデータベースは、各々の研究の関心である歴史資料や考古資料に関する情報を、研究論文や調査報告書から収集してまとめたものである。基本的に資料に関するデータベースであるから、所蔵資料データベースと同じ性質を持つ。実際、表6に示すように、博物館コアメタデータに問題なく対応付けることができる。所蔵資料データベースの対応付けを示す表5と比較すると、それぞれの情報の出典、あるいは遺跡とセットになっている場合は、その遺跡との対応が記述されることから、関連の要素が豊富である点に特徴がある。

### 2.3.2 遺跡

表2の遺跡に属するデータベースのデータ項目より、共通する性質を遺跡に関する記述要素として抽出し、これに各データ項目に対応させたものを表7に示す。遺跡に関するデータベースは、統合検索のⅡ期のシステムにおいても、DCを共通メタデータとしている。個々のデータ項目をDCに直接マッピングしているが、表6の記述要素を通して見た時のDCとの対応を表4に示している。

遺跡の記述には、数や遺跡構造といった物理的な属性が現れるため、資料のものとしての属性をDCへ対応付けることと同様な問題がある。数は博物館コアメタデータの数量と同じ意味であり、遺跡構造は形状と意味が近い。他の要素も、表4より読み取れる博物館コアメタデータと遺跡に関する要素の対応関係に違和感はない。博物館コアメタデータの要素の名称を抽象化する必要があるかもしれないが、遺跡に関する共通メタデータとして、DCより博物館コアメタデータが親和性があると言える。

### 2.3.3 事項

表2で事項としたデータベースも、DCを共通メタデータとし、個々のデータ項目を直接マッピングしている。これらのデータベースの記述対象は様々であり、遺跡の場合のように、データ項目

---

表6 資料データベースのデータ項目と博物館コアメタデータの要素との対応

メタデータ要素		棟札	中世制札(制札)	土偶	弥生石器(図面)	東国板碑(板碑)
識別番号				自治体+遺跡+土偶コード	通し番号	自治体番号, 碑所在地番号, 板碑番号
名称		資料名称, 寺社名	内容見出し, 発給者	通称名	遺跡名, 石器種別	主尊
種別	資料種別	種別			石器種別	
	内容種別					
主題						
人物	Creator	工匠				
	Contributor					
	Publisher					
	Description	現所有者	宛所, 所蔵者			
時間	Date			時期, 土器形式	時期	
	Temporal	年代	和暦年月日, 西暦年月日			年号月日, 西暦, 干支
空間		現所有者住所	国名	所在地, 緯度, 経度, 地図名	緯度, 経度	県名, 緯度, 経度
数量						
大きさ・質量		法量	法量(縦,横,厚さ), 縦横比	現存最大長		法量(全高, 現存高, 地上高, 厚さ, 上幅, 下幅)
形状		形状・形態		つくり		
材質		材質				
記述	内部記述					
	外部記述	調査方法, 現存の有無, 原位置, 残存の様態, 銘文の有無, 写真の有無, 実測図の有無, 備考	条数, 内容, 風化・変化, 釘穴・釘痕, 木目, 備考	遺跡名, 出土遺構, 所蔵者, 管理番号, 遺存部位, 実測者, 撮影, 画像, 備考, 特記事項,	石器分類	法名俗名, 造立趣旨, 装飾, 真言, 傷, 石材, 図像, 拓本, 写真, 画像, 記入者, 備考
関連		出典		文献, 報告者・著者, 報告書・論文名, 年月, 掲載紙, 刊行所	遺跡番号, 図版番号	文献番号

の全てを共通性を捉えて記述要素でくくることは出来ない。マッピングの性質をみるため、各データ項目と博物館コアメタデータの要素との関係として、表8に示し、この要素とDCとの関係を表4に示す。但し、表8には出現した要素だけを示している。

表8において、共通性のないデータ項目は記述に対応させていることが要因ではあるが、博物館コアメタデータの要素と無理なく対応している。また、物理的な属性に関するデータ項目は出現していない。したがって、DCとの対応を表す表4を見ても、種別については検討を要するが、他は無理なく対応する。

表7 遺跡データベースのデータ項目と記述要素との対応

属性	陶磁器 出土遺跡	近世窯業 遺跡	城館城下 発掘	弥生石器遺跡	東国板碑 (遺跡等)	縄文・弥生 集落遺跡
番号		窯跡番号		遺跡番号	板碑所在地番号, 自治体番号	
遺跡名称	遺跡名	窯名	遺跡名	遺跡名	遺跡名, 発見場所名称	遺跡名
遺跡種別	遺跡の種類		城館情報 (種類), 遺構種類	遺跡の種類	遺跡の性格	遺跡種別
時期	時代	操業時期	時代	存続幅, 土器形式		時期, 暦年代 土器型式
所在地	所在地, 北緯, 東経	所在地	所在地	所在地, 地形図番号	所在地, 緯度, 経度, 発見場所住所, 保管場所住所	所在地, 北緯, 東経
数		窯数				
遺跡構造		窯構造				
遺構情報		経営	城館発掘, 城下発掘, 文献記述, 現状	遺構の内容	現状, 遺構の内容	遺構
立地				立地, 標高・比高, 斜面の向き		標高
遺物情報	中国陶磁, 朝鮮陶磁, 日本陶磁	種別, 器種		石器種別		遺物
調査情報		調査	調査面積	調査者・機関, 調査年	記入者	調査原因
備考	特記事項	備考	特記事項	メモ欄		
関連文献・ 資料		窯文書, 文献番号		文献番号, 地図番号	板碑番号, 文献番号	報告書名, 副題, シリーズ名, 号数, 編著者, 発行者, 発行年月

表 8 博物館コアメタデータの要素に対応させた事項データベースのデータ項目

属性	日本荘園	古代・中世都市生活史(物価)	中世地方都市(都市)	江戸商人・職人	民俗語彙	俗信	文化財材料知識(色材)知識
識別番号	荘園コード	データコード	コード		NO		
名称	荘園名	品名・数量	都市名	商人名	見出し語	俗信	文化財名
種別		品目	区分(性格)	職種・所持株	分類 I~IV		文化財(大~小分類)
主題						動植物名, 標準和名	
時間	初見年	年月日	時期(世紀)				時代
空間	国名, 郡名, 参考市町村, 明治村字名, 史料村字名	地域	国名, 都道府県, 市町村名, 旧市町村名	居所, 現在地, 本国	旧地名	伝承地	地域
記述	領家・本家	貨幣・価格, 単価, 購入者, 売却者, 備考	備考(寺社名等), 備考(史料表記等)	備考	意味, 参照語, 備考		実験試料, 色, 下地, 最表面, 色表示なし, 備考
関連	出典, 遺文番号, 記録類, 地名辞典, 関係文献, 重複コード	史料	研究文献, 主な史料	出典	出典, 出版年	書名, 編・著者, 発行所, 発行年	著者名, 論文名, 雑誌名・文献名, 号, ページ, 年, 追記・補訂

### ③ マッピングの評価と考察

#### 3.1 DC マッピングの評価

DC へのマッピングに関しては、詳細表示において、特定の要素にデータ項目が集中することがないように、各要素に分散するような対応付けを基本とした。このような点を含め、実際のマッピングが、どのように行われているかを評価する。

所蔵資料データベースとそれ以外のデータベースについて、DC の各要素へ対応づけたデータ項目の総数をデータベース数で正規化して図 3 に示す。ここでは、例えば資料名称漢字と資料名称カナのように意味が同一のものや、緯度、経度のように対であるものであっても、データ項目として個別に設けられているものは、それぞれ計数した、但し、同一の項目を多数記述することができるよう、入力欄として項目を複数用意しているものは、一つに計数した。

図 3 において、Identifier のデータ項目数はおよそ 1 を示す。これは、表 3 および表 5~8 に見られるように、識別番号に相当するものがないデータベースがある一方、識別番号が複数のデータ項目から成るデータベースがあり、平均したデータ項目数が、約 1 であることをあらわす。また、Title がおよそ 2 であるのは、上述のように、名称が漢字とカナを持つことなどによる、

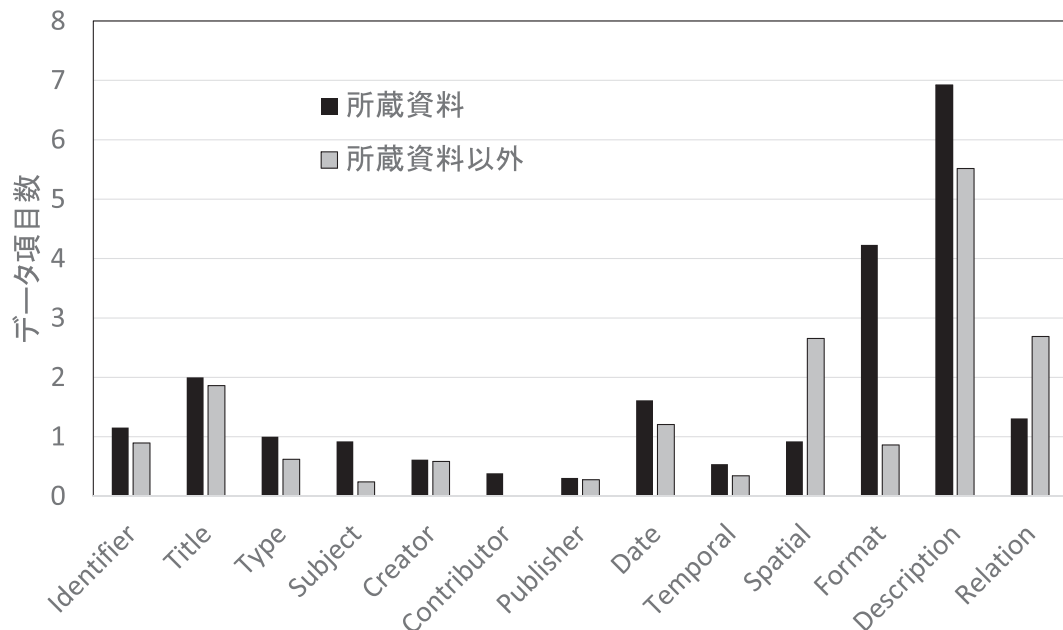


図3 DCの要素へ対応づけたデータベース当たりのデータ項目数

所蔵資料データベースは、それ以外のデータベースに比べて、Formatの項目数が大きく、物理的な属性の記述が多いことが読み取れる。一方、SpatialとRelationの項目数は、所蔵資料以外に比べて低いことが見られる。

所蔵資料と所蔵資料以外のデータベースに共通する性質として、Contributorの項目数が小さいことが挙げられる。Publisherも研究文献データベースでは、表3に見られるように多くで付与されているが、全体の項目数としては小さい。

両種類のデータベースに共通して、Descriptionの項目数が6前後と他の要素に比べて高い。これは、データ項目をDCの各要素に分散するよう対応付けを行ってはいるが、独自のデータ項目、言い換えると共通性の少ないデータ項目は、Descriptionに対応付けざるを得ないことによる。このように、データ項目がDescriptionに集中しやすいことから、詳細表示を見やすくするためには、DCの要素の意味を拡張し、できる限りDescription以外の要素に対応付ける方法が有効であることが分かる。

## 3.2 DCへのマッピングの見直し

### (1) 外部機関との連携のための共通メタデータ

DCは、統合検索システムの第I期のすべてのデータベースならびに第II期の歴博では所蔵資料以外のデータベースの共通メタデータとして、詳細表示において統一的なビューを与える役割と、他の検索システムへ検索サービスを提供する際の共通メタデータとしての役割を果たす。前者に関しては、前項で示したように、現状のマッピングが有効である。一方、後者に関し、DCの要素名がそのまま表示される形態においては、例えばFormatとして員数や法量が表示されるのは違和感



をあたえる。

実際、国立国会図書館の MDL Search において、主たる対象として図書を想定した項目名による画面には、本論で示した DC マッピングによる歴博の所蔵資料の項目の表示はそぐわない。この場合、Description にデータ項目が集中しても、DC の定義に厳密に沿って、マッピングする方が望ましいともいえる。丁寧なマッピングを行うなら、連携するシステムにおいて要素名がどのように表示されるかを意識する必要がある。

## (2) 記載日付のマッピング

表 5 および 6 では、文書などの資料中に記載された日付は、Date ではなく、Temporal に対応付けている。これは、文献 [4] で論じていることであるが、古文書には写しが存在し、これに記載される日付は、写しが作られた日付を表すものではないことから、Temporal に対応付けた。

これを踏襲して、資料中に記載されている日付は Temporal としてきた。しかし、写しではない資料も多く、写しであったとしても、上記の事情は利用者に理解されづらいであろうから、Date に対応付けることが望ましいと考えられる。

## 3.3 共通メタデータの汎用化

2.3 節において、遺跡を記述するデータベースのマッピングは、DC より博物館コアメタデータの方が親和性があり、事項に関するデータベースも、博物館コアメタデータにマッピングすることに問題がないことを述べた。研究文献データベースも、人物の中で Publisher としての役割が失われる点を除けば、博物館コアメタデータにマッピングしても不都合は生じないことが、表 4 の該当する欄の対応関係より読み取ることができる。このように、博物館コアメタデータは、資料だけでなく、人文科学のより広い研究資源の統合検索のための共通メタデータの候補となり得る。この点について、さらに考察を加える。

1.2 節で、第 II 期のシステムから、基本共通メタデータを導入したことに触れた。これは、識別番号、名称、人物、時間、空間、ならびに、種別と主題を合せてキーワードとした要素により構成している。これらの要素を選んだのは、第 I 期の W1 メタデータの中で、Who, When, Where は、意味が明瞭であって、検索の絞り込みに有用であり、DC の Title に当たる要素と種別、主題を表す項目は、資源を特徴づけ、検索の手掛かりを与え、検索結果の一覧表示にも有用と考えられることによる。ここで重要なのは、人文科学系の資源の記述に、人、時、処の 3 つの要素が重要であることと、その役割や出来事の内容は資源を探す際には直接必要とされないことにある。このように、具体的な内容を省略し抽象化した要素で構成している点が検索の手段として効果をもたらしている。

博物館コアメタデータは、基本共通メタデータと同類の抽象化され検索と一覧表示に有用な要素と、資源毎の固有の記述を統一的に詳細表示できるよう整理した要素から構成していると見ることができる。博物館コアメタデータが、人文科学の広い範囲のデータベースに対して、DC より親和性があるのは、基本共通部が抽象化されている点大きい。一方、資源毎の固有の記述に関する要素のうち、特に資料のものとしての属性を表す要素は、博物館資料の記述にチューニングして定めたものである。要素の意味と名称を抽象化するとともに、要素の共通化の可能性を探ることにより、

---

より広い研究資源を対象とする共通メタデータを構成できると考えられる。

博物館コアメタデータではDCのRightsとLanguageに相当する要素を持たせていない。これは、このメタデータを定める際に参考としたデータベースが、これらに関する項目を有していなかったことによる。その後、人間文化研究機構の在外日本関連資料調査のプロジェクトで扱った資料に関するデータベースにおいて、Languageの項目を必要とした。汎用化に向けて、共通メタデータが対象とするデータベースの範囲を十分に吟味し、そのデータ項目を広範に調査し、要素を決定してゆく必要がある。

## むすび

人間文化研究機構の統合検索システムにおいて、多数で種類が多様なデータベースを横断的に検索する上で主要な役割を果たす共通メタデータとして選択したメタデータに、歴博のデータベースのデータ項目をマッピングする上での考え方を記し、そこでの課題より汎用的な共通メタデータのあり方を考察した。

共通メタデータとして採用したDCメタデータに、実物の資料を記述したデータベースを適用することは、要素の意味の拡張が必要となって、適合性が低い。一方、博物館コアメタデータは、資料だけでなく、遺跡や種々の文献から研究の関心を抽出した事項、さらに研究文献を対象としたデータベースに対しても、そのマッピングに親和性がある。これは、記述の要素として役割や出来事の内容を問わない形で人物、時間、空間を取り出すとともに、記述対象を特徴づける要素を、共通性の高い種別、主題として取り出していることが大きいことを述べた。

人文科学系の研究対象は多種多様であり、研究の視点も様々であるから、その記述も多様なものとなる。統合検索システムは、データベースとして記述された大量の研究資源のなかから、個々の研究者がその対象を見出す手段を提供する。そこでキーとなる共通メタデータは、多様な記述をまとめて見ることができるよう抽象化され汎用的なものでなければならない。DCはその出発点であり、博物館コアメタデータは一步進んだものである。より、広範な人文科学の研究資源を対象とし得る共通メタデータを創り出してゆく上で、本論がその下地となれば幸いである。

## 謝辞

研究資源共有化システムは、多くの人との協力により構築が図られた。その中での討議は、本論をまとめる上で非常に有用であった。この構築に参画された方々に厚く感謝する。中でも国立民族学博物館の山本泰則氏とは、博物館コアメタデータの検討においても、貴重な意見交換を行うことができた。ここに深謝する。そして、本検討が可能であったのは、歴博の開設以来、進められてきたデータベースの構築があったことによる。データベースの構築に携われた方々、ならびにマッピングする際に意見をいただいた方々に深く感謝する。

---

参考文献

- [1] 安達文夫, 山本泰則, 関野 樹, 原正一郎, 柴山 守, 安永尚志, “人文科学研究データベースの統合検索システムの構築,” 画像電子学会第 36 回年次大会予稿集, R.2-1 (CD-ROM) (2008).
- [2] 山本泰則, 原正一郎, 柴山 守, 安達文夫, 合庭 惇, 安永尚志, “Dublin Core メタデータと Z39.50 にもとづく人文科学系データベースの統合検索に関する実証実験,” 情報処理学会シンポジウム論文集, vol. 2004, No.17, pp.199-205 (2004).
- [3] “DCMI Metadata Terms”, Dublin Core Metadata Initiative, Aug. 2006.  
<http://dublincore.org/documents/dcmi-terms/>
- [4] 安達文夫, 鈴木卓治, 小島道裕, 高橋一樹, “情報資源共有化のための博物館資料データベースのマッピングとその評価,” 国立歴史民俗博物館研究報告, vol. 125, pp.185-214 (2006).
- [5] “Dublin Core Metadata for Resource Discovery”, Network Working Group Request for Comments : 2413, Sep. 1998,  
<ftp://ftp.isi.edu/in-notes/rfc2413.txt>
- [6] “Guide to Best Practice : Dublin Core Version1. 1”, CIMI Consortium, Apl. 2000.  
[http://www.cimi.org/public\\_docs/meta\\_bestprac\\_v1\\_1\\_210400.pdf](http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf)
- [7] 安達文夫, “博物館資料情報の共有化のための共通メタデータへのマッピング,” 画像電子学会第 4 回画像ミュージアム研究会予稿集, 2 (2006).
- [8] 安達文夫, “人文科学情報共有化のための博物館資料情報のマッピングの検討,” 画像電子学会第 34 回年次大会予稿集, pp.179-186 (2006).
- [9] 安達文夫, 鈴木卓治, “歴史研究データベースの Dublin Core へのマッピングとその課題,” 情報処理学会研究報告, vol. 2006, No.112. 2006-CH-72, pp.47-54 (2006).
- [10] 山本泰則, 安達文夫, “博物館資料情報統合検索のためのコアメタデータ,” 情報処理学会シンポジウム論文集, vol. 2009, No.16. pp.287-294 (2009).
- [11] 山田太造, 山本泰則, 古瀬 蔵, 安達文夫, “人文科学データベース統合検索のためのメタデータとその応用,” 情報処理学会シンポジウムシリーズ, vol. 2012, No.7. pp.71-78 (2012).
- [12] 村田良二, “博物館情報の相互運用と CIDOC CRM の役割,” 情報処理学会シンポジウム論文集, vol. 2002, No.13. pp.39-42 (2002).

(国立歴史民俗博物館研究部)

(2015 年 7 月 28 日受付, 2015 年 11 月 13 日審査終了)

## **Mapping of REKIHAKU Database's Fields to Common Metadata for Integrated Retrieval**

ADACHI Fumio

The National Institutes for the Humanities established an integrated retrieval system for the over 100 diverse databases provided by its member organizations and launched the service in April 2008. This system is characterized by its intuitive usability and enhanced algorithm to allow users to retrieve any data regardless of its operation method or location. Like most other search engines, the system enables users to retrieve information matching their search terms and shows a list and detailed description of search results. In this system, however, a common metadata set plays an important role in order to retrieve data from databases developed separately and therefore organized by different attributes. A key issue of realizing the system is how to map data attributes from the different databases to this common metadata set. This paper describes a basic idea of mapping 43 databases of the National Museum of Japanese History for this integrated retrieval system.

In order to show detailed search results in an easy-to-understand way, it is essential to map data attributes in a way that some metadata searches do not make too many requests. Based on this idea, the Dublin Core (DC) Metadata Elements, adopted as a common metadata set, is not highly compatible to search engines retrieving information from databases based on descriptions of physical materials because the meanings of the elements need to be expanded. This problem has been revealed by comparing the Museum Core Metadata Elements (which were developed to share information on the collections of the Museum) with the DC Metadata Elements. The Museum Core Metadata Elements can support the mapping of databases not only of physical materials but also on historical sites, as well as various research documents and their research purposes. This is because the Museum Core Metadata Elements can be used as description elements to retrieve information on people, time, and space, regardless of content in the form of role and achievements, and because they are often the interchangeable elements of name, type, and subject that attribute specific features to the objects being described. If there is a need to use the metadata set for more diverse purposes, it will be essential to abstract the remaining elements and explore their similarities.

Key words: transversal retrieval, resource sharing, Dublin Core meta-data, museum materials, information of humanities

---